RESEARCH STATEMENT

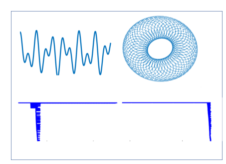
Hitesh Gakhar

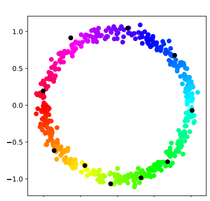
Overview

I am interested in the field of applied topology, with a focus on Topological Data Analysis (TDA). My research involves leveraging or adapting classical results from topology to develop theory or techniques for TDA and using tools from TDA for real world applications. The former includes my contributions to the theory of persistent homology, the development of sliding window persistence for quasiperiodic time series analysis, and stability results for a landmark-based non-linear dimensionality reduction scheme called Eilenberg-MacLane coordinatization. The latter comprises my most recent work on detecting methane emissions using persistence images and machine learning. Next, I give an overview of my results. In Section 1, I discuss current work and future directions. Section 2 highlights my undergraduate research experience and the questions that emerge from it.

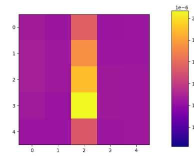
The Development of Persistent Homology: One of the most popular tools in TDA, arguably is *Persistent Homology*, which recovers and quantifies topological features of data. The success of persistence can be attributed to strong theoretical foundations and efficient algorithmic implementations. My main contribution to persistent homology is the development of two versions of the classical Künneth formula (a result in algebraic topology relating the homology of spaces to that of their products) for persistent homology [5]. The proofs use homological algebra and are presented in a generalized setting, that is, when the inclusion maps in filtered spaces are replaced by continuous maps between topological spaces. Leveraging these formulae, I also developed novel methods for algorithmic and abstract computations of persistent homology. There have been applications of these results to time series analysis and physical chemistry. My other contribution includes an algorithm to compute persistent lifts to $\mathbb{Z}_{p^k}$-coefficients of $\mathbb{Z}_p$-persistent (co)homology classes, since persistence can only be computed for coefficients in a (typically finite) field [8].

Dynamical Systems and Time Series Analysis: When a time series results from a dynamical system observation, *sliding window* (or *time delay*) *embeddings* can be used to reconstruct the topology of the traversed state space. This is a consequence of Takens' theorem [14]. In [6], we develop a general theory for using sliding window embeddings and persistent homology to detect quasiperiodic signals, which are observations of dense trajectories in toroidal dynamical systems. The results include convergence theorems both at the level of embeddings and persistent homology using Fourier polynomials of the quasiperiodic function, a theorem describing the structure of quasiperiodic sliding window embeddings, lower bounds on the persistent homology of the embeddings based on the Fourier coefficients, guidance on choosing good embedding parameters, and a demonstration to quasiperiodicity detection in music samples. The figure on the right shows a quasiperiodic signal, its toroidal sliding window embedding and persistence.



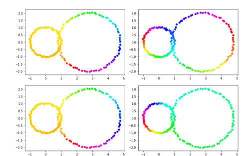


Although persistent homology is informative about the topology of data sets, one can further use it to perform nonlinear dimensionality reduction, for example on the point cloud obtained from sliding window embeddings, for the purposes of visualization and reconstruction of the underlying dynamics. This recoordinatization process, called *Eilenberg-MacLane coordinatization*, is a collection of algorithms that use persistent cohomology classes on a landmark set to construct maps from point clouds into various Eilenberg-MacLane spaces, like the circle [2, 12], projective spaces [11], or lens spaces [13]. The figure on the left shows a noisy circle (with landmarks in black) reparametrized using circular coordinates. For this method of coordinatization to be reliable, one needs stability with respect to perturbation of the landmark set, which we prove in [3, 4].



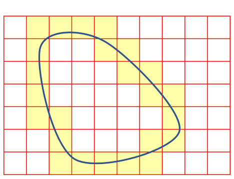
Methane Detection using Topology: The global warming potential of methane is much larger than carbon dioxide and is estimated to be responsible for a significant fraction of the temperature increases we experience. This makes measuring methane emissions an important task. In this work, we show how topology and machine learning can be used to detect emission hotspots from observed methane concentrations. The topological structure of sublevel sets of the observation function defined on 5×5 patches of data (example shown on the right) is captured and summarized using Persistence Images. These images along with time delay embeddings of a selection of patches is fed to a Convolutional Neural Network. Our network has a training set accuracy of 94.12% and a validation set accuracy of 91.3%.

1. Current & Future Work

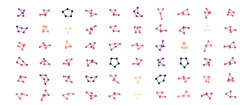


1.1. **Geometrically Independent Circular Coordinates.** The Sparse Circular Coordinates algorithm is a topological dimensionality reduction algorithm, that turns a 1-dimensional persistent cohomology class on a data set $X \subset M$ into a family of circle valued functions on X. If the data has multiple 1-dimensional features, their corresponding circular coordinates can be used to reparameterize the dataset in order to capture its essential circular geometry. However, the choice of cocycles (to generate circular coordinates) is extremely important. The cocycles picked by the algorithm often do not correspond to minimal cycles in the Rips filtration. In [1] we tackle this using a change-of-basis that is motivated by the need to have cocycle generators that are, in some sense, as "geometrically independent" or "orthogonal" as possible. The experimental methodology involved writing an inner product matrix of the harmonic cocycles and leveraging that to perform a Gram-Schmidt type orthogonalization process for integer cocycles, called the LLL basis reduction algorithm [9]. A demonstration of this is presented in the figure on the right. In the top row, we see circular coordinates given by DREiMac. In the bottom row, we see circular coordinates built from the orthogonalized cocycles. *This project started at the Mathematics Research Communities 2022 (Data Science at the Crossroads of Analysis, Geometry, and Topology) and is in collaboration with Johnathan Bush, Jose Perea, Tatum Rask, Nikolas Schonsheck, Luis Scoccola, and Ling Zhou.*

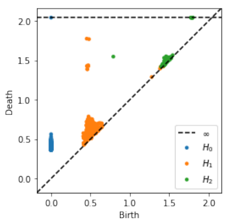
1.2. **Manifold Reconstruction and Dimensionality Reduction.** Manifold reconstruction is a process of building a representation of an unknown manifold from its sample, or determining certain topological properties. In some applications, it is done via creating a triangulation of the manifold. Unfortunately, this problem of building triangulations becomes computationally intractable as the dimension of manifold increases.



However, if one is interested in the 'shape' of the manifold, then understanding its homology groups is often sufficient. If the point cloud sampled from M is sufficiently dense, then the union of balls centered at the point cloud deformation retracts to M [10]. Consequently, the homology groups of M can be computed from the nerve of this neighborhood (also known as the Čech complex), or using Vietoris-Rips complex via interleavings. However, this approach is computationally expensive when trying to find high dimensional homology groups. We are working on building homotopy preserving cubical approximations to a manifold and cubical reconstructions from its samples with sufficient density. Since the dimension of this constructed complex depends on that of the ambient dimension, we are also working on a dimensionality reduction technique so the (reduced) complex better reflects the intrinsic dimension of the manifold. The advantage of this technique is that it can be used for efficient computations of the homology groups of higher dimensional manifolds. Currently, we have results for planar curves and some progress on the general cases. *This collaboration started at the University of Oklahoma with Miroslav Kramar and Reza Niazi.*



1.3. **Topology of Equilateral Pentagonal Linkages.** The space $\mathbb{M}$ of planar pentagonal linkages of unit length is the set where two pentagons are considered equivalent if they differ by a rotation or a translation (but not reflection). Spaces like $\mathbb{M}$ serve as toy models for confirmation spaces of molecules and their energy landscapes. In the figure on the right, we see a sample collection of some pentagons along with their energy (computed by oriented area) represented by their color.

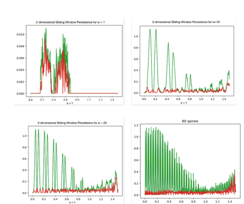


What makes $\mathbb{M}$ interesting is the fact that even though it can be embedded in $\mathbb{R}^6$, topologically it is a 2-dimensional orientable manifold of genus 4, as proved in [7] as an application of coordinate geometry. In the figure on the left, we show the Rips persistent diagram that verifies the shape of $\mathbb{M}$. The long term goal of this work is to develop topological tools to understand spaces like $\mathbb{M}$, and the persistent homology of the sublevel set filtration induced by the energy function. Using this information, one could find a path along $\mathbb{M}$ from one stable configuration (local minima on the energy surface) to another while minimizing energy costs. *This project started at the Mathematics Research Communities 2022 (Data Science at the Crossroads of Analysis, Geometry, and Topology) and is in collaboration with Johnathan Bush, Alex Elchesen, Araceli Guzmán Tristán, Iryna Hartsock, Barbara Mahler, Francisco Martinez, and Jose Perea.*

1.4. **Vietoris-Rips Persistence of a Torus.** In [5], we showed an application of the categorical Künneth theorem to compute the Vietoris-Rips persistent homology of the torus equipped with the maximum metric. However, a more natural choice is the product metric $d_{\mathbb{T}^2}(p,q) = \sqrt{d^2_{S^1}(p_1,q_1) + d^2_{S^1}(p_2,q_2)}$ on $\mathbb{T}^2$. Recently, I started investigating the Vietoris-Rips persistence of $n \times n$ uniformly spaced points in the torus $\mathbb{T}^2$ viewed as a product of unit perimeter circles. My approach was to start experimentally to track persistence patterns as the number of points increase. Some trivial conclusions were that the Rips complex would always start out as n^2 different components, which turns into a wedge of n^2+1 circles at $\frac{1}{n}$, which then becomes the torus at $\frac{1}{n}\sqrt{2}$ which stays the same way topologically until the scale hits $\frac{1}{n}k$ where $n = 3k-2, 3k-1, 3k$. The patterns after this become harder to spot, for example, for different n, $\mathbb{T}^2_n$ turns into either a wedge of 2-spheres, 3-spheres, 4-spheres or all. The only common observation is the death of the 1-dimensional persistence. My long term plan is to investigate the persistence further to get an idea of what the spaces could be and then use theoretical justification to confirm that.

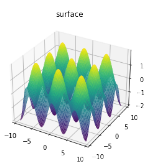
2. Undergraduate Research Projects

I have served as a research mentor to two undergraduate students at the University of Oklahoma. Both students learned the basics of topological data analysis and performed experiments in the realm of topological time series analysis. One of them submitted their Honors Thesis in May 2022; the other will submit it in December 2022.



Observations of typical circular or toroidal dynamical systems are recurrent, with periodicity and quasiperiodicity being the two prominent types. In [15], the authors found a class of functions that could be viewed as generic observations of a dynamical system on sphere (among other popular manifolds). In such a system, a point on this dynamical system moves from the north pole to the south pole winding around the sphere. Specifically, the trajectory would appear similar to a thread being wrapped around a ball going from pole to pole. My student *Mark Ramirez* varied the choice of *winding coefficient*, that is how tightly the trajectory of the point winds around the sphere, and investigated its effect on sliding window persistence and the choice of delay parameter τ that minimizes topological distortion. As shown on the right, for different winding coefficients (represented by different images), Mark computed the sliding window persistence in dimension 2 for hundreds of values of τ and tracked the discrepancy between the top two persistence values (plotted by the green and the curves). A τ value is a good choice if the curves are separated sufficiently. We observe that the tighter the winding gets, the more frequent the good choices of τ get. His experiments lead to a few more questions that would be suitable for future undergraduate projects:

- How are the good choice delay parameters τ distributed?
- How do our observations change for dynamical systems on higher-dimensional spheres?
- How do our observations change when the sphere is distorted into an ellipsoid or an hour-glass?



While there has been development for sliding window persistence of real-valued and vector-valued functions, there has been no work for detecting recurrence in surfaces, i.e. graphs of multivariable functions (e.g., the surface on the right). The theoretical development in this problem will use tools from multivariate Fourier analysis, as in [6]. Applications of this could be in the field of structural chemistry, either in detecting quasicrystallic structure or quantifying recurrence in potential functions. My current student *Anishka Peter* is studying the first examples of sliding window persistence of multivariable functions. Anishka will look at two strategies of quantifying recurrence. In the first, we will slice the multivariable domain, classify recurrence along each one-dimensional slice, and then use it to argue for multivariate recurrence. In the second, we will change the notion of the sliding window to a “sliding box” or a multivariable window to generate the point cloud necessary for persistence. Some questions that we are interested in are:

- How do the two strategies differ in results and computational performance?
- In the first strategy, how does one choose the correct way to slice the domain?
- In the second strategy, what are the optimal choices of the vector-valued parameter time-delay $\vec{\tau}$?

References

[1] J. Bush, H. Gakhar, J. Perea, T. Rask, N. Schonsheck, L. Scoccola, and L. Zhou. Geometrically independent circular coordinates via lattice reduction. *in preparation*, 2022.

[2] V. De Silva, D. Morozov, and M. Vejdemo-Johansson. Persistent cohomology and circular coordinates. *Discrete & Computational Geometry*, 45(4):737–759, 2011.

[3] H. Gakhar, J. L. Mike, J. A. Perea, and L. Polanco. Stability of multiscale K(G,1) coordinates. *In preparation*, 2022.

[4] H. Gakhar, J. L. Mike, J. A. Perea, and L. Polanco. Stability of sparse circular coordinates. *In preparation*, 2022.

[5] H. Gakhar and J. A. Perea. Künneth formulae in persistent homology. *arXiv preprint arXiv:1910.05656*, 2019.

[6] H. Gakhar and J. A. Perea. Sliding window persistence of quasiperiodic functions. *arXiv preprint arXiv:2103.04540*, 2021.

[7] T. F. Havel. Some examples of the use of distances as coordinates for euclidean geometry. *Journal of Symbolic Computation*, 11(5-6):579–593, 1991.

[8] J. Melby, H. Gakhar, and J. A. Perea. Persistent lifts to $p-$primary coefficients. *In preparation.*

[9] P. Q. Nguyen and B. Vallée. *The LLL algorithm*. Springer, 2010.

[10] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1-3):419–441, 2008.

[11] J. A. Perea. Multiscale projective coordinates via persistent cohomology of sparse filtrations. *Discrete & Computational Geometry*, 59(1):175–225, 2018.

[12] J. A. Perea. Sparse circular coordinates via principal $\mathbb{Z}$-bundles. In *Topological Data Analysis*, pages 435–458. Springer, 2020.

[13] L. Polanco and J. A. Perea. Coordinatizing data with lens spaces and persistent cohomology. *arXiv preprint arXiv:1905.00350*, 2019.

[14] F. Takens. Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980*, pages 366–381. Springer, 1981.

[15] B. Xu, C. J. Tralie, A. Antia, M. Lin, and J. A. Perea. Twisty takens: A geometric characterization of good observations on dense trajectories. *Journal of Applied and Computational Topology*, 3(4):285–313, 2019.